Seasonal statistical-dynamical prediction of the North Atlantic Oscillation by probabilistic post-processing

André Düsterhus[1,2]

[1]Institute of Oceanography, Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Germany
[2]ICARUS, Department of Geography, Maynooth University, Ireland

Correspondence: André Düsterhus (andre.duesterhus@mu.ie)

Abstract. Dynamical models of various centres have shown in recent years seasonal prediction skill of the North Atlantic Oscillation (NAO). By filtering the ensemble members on the basis of statistical predictors, known as subsampling, it is possible to achieve even higher prediction skill. In this study the aim is to design a generalisation of the subsampling approach and establish it as a post-processing procedure.

5   Instead of selecting discrete ensemble members for each year, as the subsampling approach does, the distributions of ensembles and statistical predictors are combined to create a probabilistic prediction of the winter NAO. By comparing the combined statistical-dynamical prediction with the predictions of its single components, it can be shown that it achieves similar results to the statistical prediction. At the same time it can be shown, that unlike the statistical prediction the combined prediction has less years where it performs worse than the dynamical prediction.

10   By applying the gained distributions to other meteorological variables, like geopotential height, precipitation and surface temperature it can be shown that evaluating prediction skill depends highly on the chosen metric. Besides the common anomaly correlation (ACC) this study also presents a score basing on the Earth Mover's Distance (EMD), which is designed to evaluate skill of probabilistic predictions. It shows that by evaluating the predictions separately compared to applying a metric on all years at the same time, like correlation based metrics, leads to different interpretations of the analysis.

1   Introduction

Seasonal prediction of the North Atlantic Oscillation (NAO) is a challenge. During the year the NAO describe a high portion of the explained variability of the pressure field over the North Atlantic region and with it has a high influence on European weather. While the Winter-NAO (WNAO) is a dominant factor for changes in the storm tracks over the North Atlantic (Hurrell

20   (1995)), the Summer-NAO (SNAO) is associated with precipitation and temperature differences between Scandinavia and the Mediterranean (Folland et al. (2009)).

Current seasonal prediction systems have demonstrated limited skill for the WNAO (Butler et al. (2016)). Dobrynin et al. (2018) has shown that by combining statistical with dynamical prediction, a much higher significant skill is achievable. The paper applies an ensemble subsampling algorithm, which bases on selection of ensemble members on the basis of statistical predictors. The selected ensembles are then used to create a new sub-selected ensemble mean, which has for the NAO index, but also for many other variables and regions, a better prediction skill than the ensemble mean of all ensemble members.

Statistical-dynamical predictions basing on different strategies are common in many fields in geoscience. Gleeson (1970) developed a framework for the dynamical evolution of statistical distributions in phase space with applications to meteorological fields. Vecchi et al. (2011) apply a combined statistical-dynamical approach by using an emulator basing on dynamical forecasts to create seasonal hurricane prediction. Roulston and Smith (2003) developed a "best member" concept, which use verification statistics to dress an dynamical ensemble prediction.

The focus of this paper is to implement the sub-sampling algorithm as a probabilistic post-processing procedure, demonstrated at the hand of the seasonal prediction of the WNAO. It interprets ensemble members and the statistical predictors as values with uncertainties and combines them to a new probabilistic forecast. This approach allows us to evaluate a prediction skill not only for a long time series, but for each individual year. We use for this a newly developed skill score, the 1D-continuous-EMD score, basing on the Earth Mover's Distance (EMD). We also use the probabilistic information of the prediction to create a weighted mean of the ensemble members, which creates a better hindcast skill for important meteorological variables like surface temperature and precipitation.

2 Data and Model

To demonstrate the procedure we use the seasonal prediction system based on the MPI-ESM (Dobrynin et al., 2018) with a model resolution of T63/L95 in the atmosphere and T0.4/L40 in the ocean (also known as mixed resolution, MR). As described by Baehr et al. (2015), we initialise in each November between 1982 and 2017 a 30 ensemble member hindcast from an assimilation run based on assimilated reanalysis/observations in the atmospheric, oceanic and sea-ice component.

In these hindcasts the NAO is calculated by an EOF analysis (Glowienka-Hense, 1990). For the WNAO we calculate the mean sea level pressure field for December, January and February and calculate the EOF of the North Atlantic sector limited by $20°-80°$N and $70°$W$-40°$E.

3 Methodology

3.1 Seasonal prediction of the WNAO

The seasonal prediction of the WNAO for the period of 1982 to 2017 is shown in Figure 1. Every dot represents one WNAO value of one ensemble member, which has also available the full meteorological and oceanographical fields during the associated winter period. These hindcast predictions for the WNAO have a large spread, covering the range of the observations given by the reanalysis, but do not give indication for a specific NAO value 2 to 4 month ahead. The skill is generally evaluated as a

correlation skill, which have been shown in recent years to have hindcast skill for several different prediction systems (Butler et al., 2016).

3.2 Statistical-dynamical prediction

Our approach will be applied to every single year independently. As an example we choose the year 2010, which shows an extreme negative WNAO value. The first step is to generate one probability density function (pdf) for each ensemble member prediction (e_i) of the WNAO value, which is generated by a 2000 member bootstrap of the EOF fields (Wang et al., 2014). To create from these predictions a pdf for all ensemble members ($\mathcal{E}$), mixture modelling (Schölzel and Hense, 2011) at discrete NAO index values is applied:

$$\mathcal{E}(v) = \sum_{i \in \mathcal{I}} e_i(v) \tag{1}$$

Here, v corresponds to each value of the discretised NAO values, and $\mathcal{I}$ to the indices of the ensemble members. The chosen resolution for the discretised NAO values is 0.01 and after creating the sum of all single member pdfs, the overall pdf $\mathcal{E}$ is normalised. For 2010 the results are shown in Figure 2. As expected from Figure 1, the dynamical model prediction has a very broad pdf equating to a low signal.

To sharpen the prediction we introduce literature backed physical statistical predictors. As predictors (p_i) we use those defined by Dobrynin et al. (2018): sea-surface temperature in the northern hemisphere, Arctic sea-ice volume, Siberian snow cover and stratospheric temperature in 100 hPa. All predictors and their influence on the WNAO have been discussed in the paper. For the physical validity of a prediction the selection of the correct predictors is essential and have to be adapted to any newly analysed phenomena individually. Each predictor predicts statistically from the climatic state taken from the ERA interim reanalysis (Dee et al., 2011) before the model initialisation a WNAO value in the following winter.

We treat the predictors p_i like the ensemble members before and apply an empirical mixture modelling. For the year 2010 the results are shown in Figure 3. Due to the limited number of predictors compared to the ensemble members, and in the shown case also due to their alignment, the resulting overall predictor pdf ($\mathcal{P}$) is much sharper than the overall model prediction.

To create a combined prediction the two pdfs ($\mathcal{E}$ and $\mathcal{P}$) are after normalisation multiplied at each of the discretised NAO values:

$$\mathcal{M}(v) = \mathcal{E}(v) \cdot \mathcal{P}(v). \tag{2}$$

After another normalisation the final combined prediction $\mathcal{M}$ creates the statistical-dynamical prediction for the seasonal NAO prediction in the specific year. The result for the year 2010 is shown in Figure 4. The pdf of the combined prediction is close to the one of the predictor predictions, but shows differences where there is additional information from the model prediction. Therefore, the combined prediction shows a clearer signal as the model prediction, which does not give any indication of a specific NAO value at all.

3.3 NAO evaluation

To evaluate the performance of the three different predictions ($\mathcal{E}$, $\mathcal{P}$ and $\mathcal{M}$) and compare the predictions with the observation,

85 we use the Earth Mover's Distance (Rubner et al. (2001)) to define the 1D-continuous-EMD-score. The one-dimensional EMD (Düsterhus and Hense (2012)) can be derived by

$$D(f,g) = \frac{1}{N} \sum_{i=1}^{N} |F(v_i) - G(v_i)|, \tag{3}$$

where f and g are two pdfs and F and G the associated cdfs.

The 1D-continuous-EMD-score is then defined by comparing the pdfs of the model prediction ($\mathcal{M}$), the observations ($\mathcal{O}$)

90 and the climatology ($\mathcal{C}$). It is calculated by:

$$q(\mathcal{M},\mathcal{O}) = 1 - \frac{D(\mathcal{M},\mathcal{O})}{D(\mathcal{C},\mathcal{O})} \tag{4}$$

In case of a perfect prediction the score becomes 1, a model prediction equal to a climatology 0 and negative for a worse prediction than the climatology. Since the NAO index is normalised for mean and standard deviation, we use as climatology a standard normal distribution $\mathcal{N}(0,1)$. To estimate uncertainties, we use 500 randomly selected uniformly distributed weightings

95 of the ensemble members between 1 and 0 and create with those a pdf for the scores.

3.4 Variable field evaluation

To estimate the post-processed variable field, we calculate a weighted mean of the meteorological variable fields. The weighting coefficients are estimated by weighting the model prediction $\mathcal{M}$ with each of the pdfs (e_i) of the ensemble members:

$$c_i = \sum_v e_i(v) \cdot \mathcal{M}(v) \tag{5}$$

100 This generates a coefficient for each ensemble member and the weighted mean of the atmospheric field then generates the model prediction for the field.

For evaluating the meteorological variable fields we apply two different strategies. The first is the Anomaly Correlation Coefficient (ACC), a common measure of skill in seasonal predictions. The second approach is using the 1D-continuous-EMD score at every grid point. As a climatology all observational values for the investigated time frame are chosen. The observation

105 in each year is a single value with 100% as a weight. In case of the weights for the ensemble member, each value of the variable at the grid point gets weighted with the relative weight c_i given by the three different prediction. By this approach it is possible to calculate the 1D-continuous-EMD score for each of the three different predictions. In section 4.2.2 the relative positioning between two predictions is shown. Significances are here determined by DelSole and Tippett (2016), which determines the skill significances by comparisons to random walks.

Table 1. Count of years of relative positioning of the three different predictions using the median of the 1D-continuous-EMD score.

rank	WNAO		
	model	predictor	combination
1	5	17	14
2	5	10	21
3	26	9	1

4 Results

4.1 Evaluating the seasonal NAO prediction

In a next step we evaluate the yearly performance of the WNAO prediction of the three different predictions ($\mathcal{E}$, $\mathcal{P}$ and $\mathcal{M}$) with the 1D-continuous-EMD score. Figure 5 shows that the results of the combined ($\mathcal{M}$) and the predictor ($\mathcal{P}$) prediction are clearly better performing than the the model results ($\mathcal{E}$). In most years, the combined and the predictor prediction demonstrates skill compared to a climatological prediction over the whole uncertainty range. The model prediction has less variability over the years in skill than the other two predictions and only in a few years is able to reach the average skill of the combined prediction. The median and interquartile range of the summed up prediction skill for all evaluated years for the combined prediction ($0.39\ [0.21; 0.60]$), is higher compared to the model ($0.12\ [0.01; 0.22]$) and predictor ($0.37\ [0.17; 0.56]$) prediction. There is only one year (2003), with a strong discrepancy of the combined and predictor prediction and a clearly negative score.

To better evaluate the performance of each prediction in contrast towards the other predictions, we determine the relative ranking of the median of each prediction in each year. The rankings are counted for the whole hindcast period and the results displayed in Table 1. The model prediction has only in a few years a better prediction skill than the other two predictions. In the majority of the years its prediction skill is lower than both other prediction. Looking at the best prediction for each year, the predictor and combined prediction is on equal terms. Nevertheless, the combined prediction is much more unlikely to be the worst of the three predictions in a year, while the predictor prediction takes much more often the third rank. These results show that the combined prediction is closer to the statistical rather than the dynamical prediction, but in case one of the two is better than the combined prediction, the latter beats in nearly all cases the one remaining. As such it offers a smoothing of the prediction skill, preventing many worse predictions.

4.2 Analysis of atmospheric variable fields

4.2.1 Climatological analysis

In the following we investigate three different atmospheric variable fields: surface temperature, total precipitation and 500 hPa geopotential height. In figure 6 the results are shown for the winter (DJF) season with the Anomaly Correlation Coefficient (ACC).

For the winter surface temperature, the main areas of significant hindcast skill of the combined prediction can be found over large parts of the North Atlantic and in a band reaching from Northern France to Eastern Europe, sparing North Scandinavia and the Mediterranean. These results are comparable to those shown by Dobrynin et al. (2018). Comparing it to the model prediction shows that the main significant increase in skill can be found over Western Europe with a general non-significant increase over the whole continent. Also some significant increase of prediction skill can be found in the Labrador Sea, while a significant decrease is located over Greenland. The comparison to the predictor prediction shows that the differences are hardly existing. The areas shown as significant have to be assumed as random and an artefact of the bootstrapping approach.

The total precipitation has significant positive hindcast skill North of the British Isles, east of the Baltic Sea, in the Mediterranean and between the Canaries and the Azores. Towards the model prediction the area east of the Baltic sea and the Mediterranean is significantly increased, while again compared to the predictor prediction not much change is detectable. Finally for the geopotential height, the hindcast skill for the combined prediction is found in areas over Iberian peninsula, between the Canaries and the Azores and between the British Islands and Greenland. Compared to the model prediction, some increase can be found over southern Scandinavia and east of Greenland. In the comparison to the predictor prediction areas over Greenland and the British Isles are significantly decreased in their hindcast skill. This can be explained by the conditioning of the predictor on the NAO directly, while the dynamic component of the statistical dynamical prediction decreases the skill in the main inflence areas of the NAO.

The analysis shows that there exist changes between the model and the combined prediction. Generally, the hindcast skill of the combined prediction is very close to the one of the predictor prediction.

4.2.2 Analysing single years

In a final step, the same atmospheric fields are compared with the 1D-continuous-EMD score. Again the analysis shows the relative positioning of two predictions. For the surface temperature in Winter the difference between the model and combined prediction is only significant in small patches distributed over the North Atlantic. Generally, no clear patterns can be identified. Especially the significant areas determined by the ACC before do not show any significance with this score. The significant area in the ACC over western Europe has some increased values in favour of the combined predictions, but is not significant. In the comparison between the predictor and the model prediction the increases and decreases are consistent, which shows that the reasons for the differences between the combined and the model prediction can be explained by the statistical model. Better hindcast skill for the combined prediction compared to the predictor prediction can be identified in the west of the Mediterranean.

For the total precipitation the only significant change is a stretch north of Scandinavia. Also for the other comparisons for this variable the changes are small. This is different for the geopotential height, where large areas in the North-East Atlantic

165 and north of Scandinavia are significantly better represented in the combined prediction rather than the model prediction. Both areas are not identified in the equivalent comparison with the anomaly correlation. The comparison of the predictor prediction compared to the model prediction shows very similar patterns. The last comparison shows that the combination has areas between the Canaries and the Azores where it is significantly higher, while in large areas of Western Europe it has more often a better skill, but is not significantly better.

170 This analysis shows that the three prediction do not have in all cases a clear relative ranking towards each other. Generally the results are very patchy and apart from north of Scandinavia no consistency can be seen.

5 Discussion and Conclusion

This paper shows a post-processing procedure, generalising the newly established subsampling procedure by Dobrynin et al. (2018). By selecting not only single ensemble members, but utilising their uncertainty ranges, a much better understanding on

175 the reason for its success is possible. As seen in section 3.2 the better prediction skill for the NAO by the combination of the statistical and dynamical model compared to the unprocessed dynamical prediction results from the sharper prediction of the statistical prediction. As by construction the different statistical predictions are highly connected towards the target value, in this case NAO, the predictor driven predictions result in higher skill. Furthermore, advantages of using this post-processing approach compared to the pure subsampling is the availability of non-parametric uncertainties for the predictions and the

180 possibility of weighting the different ensemble members for the analysis of variable fields with unequal weights. As such especially outliers can therefore be much better handled, without giving them a too high weight within the analysis.

Compared to the predictor prediction, the combined prediction achieves similar results for the NAO prediction. In a three-way comparison together with the dynamical prediction we have shown that it is not generally showing more skill than the predictor prediction, but it observes less negative outliers in skill. Nevertheless, in case of the atmospheric variable predictions,

185 the predictor prediction is not entirely a statistical prediction. The construction of weighting the ensemble members, leads to statistical dynamical prediction as well, where the weight of the dynamical model is less pronounced. As such, the skill between the two dynamical-statistical predictions is more similar in this case than the NAO prediction itself. We have seen that the two used scores show the hindcast skill of the different forecasts in different ways. The 1D-continuous-EMD score allows to effectively evaluate the skill of two probabilistic results, like observations and predictions. The score has similar characteristics

190 like the RMSE in cases of undetected trends, different variability of different forecasts or a bias. In case of this study it is noted that the combined prediction has generally a larger variability as the model prediction, resulting from the sharpened predictions of the combined approach. An important advantage compared to skill measures like correlation or RMSE is the possibility to evaluate non-normal distributed variables. Also unlike the correlation an evaluation of the skill is possible to be decomposed for every single year in a consistent way. Not shown here is that a detrending and normalisation of the variables would lead to

195 a score, which is in characteristics in many ways similar to the correlation, without losing its advantages.
As each year is compared to the climatology, in case of an NAO value close to zero, a prediction in one direction can have a big influence and results in substantive negative scores. To prevent this other references, like uniform distributions over the whole measurement range, can be an appropriate alternative. When a harder penalty for miss-predictions is preferred, a squaring within the sum of equation 3 leads to a characteristics of the Continuous Ranked Probability Score (CRPS) for observations with uncertainties (Thorarinsdottir et al. (2013)).

By evaluating the skill on a yearly basis and taking a look at the relative positioning the approach allow for a paradigm change as also described by DelSole and Tippett (2016). By counting the years in which one prediction is better than another, a single outlier cannot anymore drive the whole verification result as it can do for correlations or RMSE. It also answers a typical question in forecast verification in a much more appropriate way: How sure can we be that a single prediction is better than another? The used evaluation procedure presented here is able to quantify this answer for non-parametric predictions.

The anomaly correlation (ACC) is well used in literature and its main disadvantage ist the assumption of a normal distribution. We have seen that there are considerable differences when all years are evaluated at the same time, like it is done in a correlation based score, or the evaluation bases on evaluating single years. Correlations can be misleading and show skill where there is not necessary a good argument for it as it is prone to outliers. These discussions are well known when correlation like measures are compared with distance-like measures, like the RMSE. Further progress in the creation of appropriate skill evaluation is therefore necessary. It is noted, that while we show in this analysis only the results for the winter season, the results for the summer season are comparable.

The methodology and verification techniques shown in this analysis are widely applicable within predictions of many different phenomena. This is especially valid in case of non-parametric datasets like in the analysis of extremes. The statistical-dynamical approach as illustrated here delivers consistent improved results compared to one of its components. Seen as a post-processing step, it forms a useful step to condition predictions on a physical basis in order to reduce noise and intensify the signal. Using non-parametric approaches in the analysis offers a more appropriate path to verify predictions in general.

Author contributions. A. D. developed the methodology, designed the study, interpreted the results and wrote the manuscript.

Competing interests. There are no competing interest by the author.

Acknowledgements. The author would like to thank Johanna Baehr and Mikhail Dobrynin for the fruitful discussions. AD was supported by the University Hamburg's Cluster of Excellence Integrated Climate System Analysis and Prediction (CliSAP).

References

Baehr, J., Fröhlich, K., Botzet, M., Domeisen, D. I. V., Kornblueh, L., Notz, D., Piontek, R., Pohlmann, H., Tietsche, S., and Müller, W. A.: The prediction of surface temperature in the new seasonal prediction system based on the MPI-ESM coupled climate model, Climate Dynamics, 44, 2723–2735, 2015.

Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez, A., Déqué, M., Domeisen, D. I. V., Fröhlich, K., Hendon, H., Imada, Y., Ishii, M., Iza, M., Karpechko, A. Y., Kumar, A., MacLachlan, C., Merryfield, W. J., Müller, W. A., O'Neill, A., Scaife, A. A., Scinocca, J., Sigmond, M., Stockdale, T. N., and Yasuda, T.: The Climate-system Historical Forecast Project: do stratosphere-resolving models make better seasonal climate predictions in boreal winter?, Quarterly Journal of the Royal Meteorological Society, 142, 1413–1427, 2016.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J., Park, B., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

DelSole, T. and Tippett, M. K.: Forecast Comparison Based on Random Walks, Monthly Weather Review, 144, 615–626, 2016.

Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F., Düsterhus, A., Fröhlich, K., Pohlmann, H., and Baehr, J.: Improved teleconnection-based seasonal predictions of boreal winter, Geophysical Research Letters, 45, 3605–3614, 2018.

Düsterhus, A. and Hense, A.: Advanced information criterion for environmental data quality assurance, Advances in Science and Research, 8, 99–104, 2012.

Folland, C. K., Knight, J., Linderholm, H. W., Fereday, D., Ineson, S., and Hurrell, J. W.: The Summer North Atlantic Oscillation: Past, Present, and Future, Journal of Climate, 22, 1082–1103, 2009.

Gleeson, T. A.: Statistical-Dynamical Predictions, Journal of Applied Meteorology, 9, 333–334, 1970.

Glowienka-Hense, R.: The North Atlantic Oscillation in the Atlantic-European SLP, Tellus, 42, 497–507, 1990.

Hurrell, J. W.: Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation, Science, 269, 1995.

Roulston, M. S. and Smith, L. A.: Combining dynamical and statistical ensembles, Tellus A: Dynamic Meteorology and Oceanography, 55, 16–30, https://doi.org/10.3402/tellusa.v55i1.12082, 2003.

Rubner, Y., Puzicha, J., Tomasi, C., and Buhmann, J. M.: Empirical Evaluation of Dissimilarity Measures for Color and Texture, Computer Vision and Image Understanding, 84, 25–43, 2001.

Schölzel, C. and Hense, A.: Probabilistic assessment of regional climate change in Southwest Germany by ensemble dressing, Climate Dynamics, 36, 2003–2014, 2011.

Thorarinsdottir, T. L., Gneiting, T., and Gissibl, N.: Using Proper Divergence Functions to Evaluate Climate Models, SIAM/ASA J. Uncertainty Quantification, 1, 522–534, 2013.

Vecchi, G. A., Zhao, M., Wang, H., Villarini, G., Rosati, A., Kumar, A., Held, I. M., and Gudgel, R.: Statistical–Dynamical Predictions of Seasonal North Atlantic Hurricane Activity, Monthly Weather Review, 139, 1070–1082, https://doi.org/10.1175/2010MWR3499.1, 2011.

Wang, Y.-H., Magnusdottir, G., Stern, H., Tian, X., and Yu, Y.: Uncertainty Estimates of the EOF-Derived North Atlantic Oscillation, Journal of Climate, 27, 1290–1301, 2014.
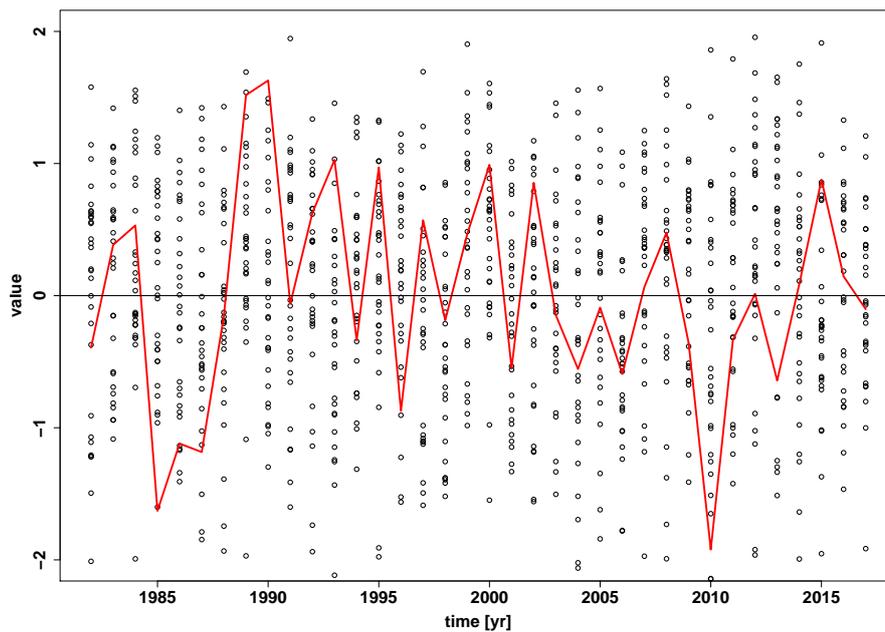
Figure 1. Seasonal prediction of the NAO. Single models (black) initialised in November predicting the DJF-NAO (red).
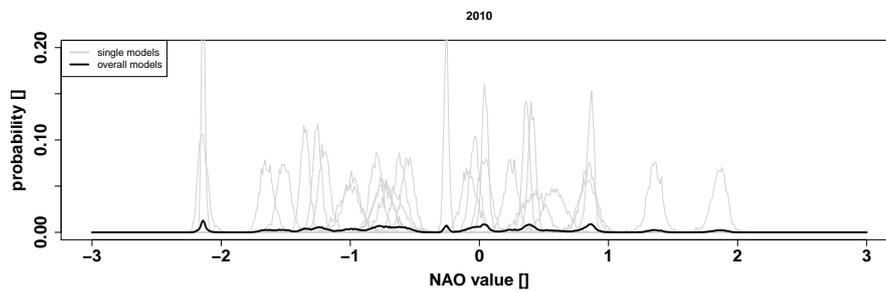
Figure 2. Model prediction of the NAO for 2010. Single models (grey) as pdf of their bootstrapped uncertainties. From this the overall model prediction (black) is created by empirical mixture modelling.
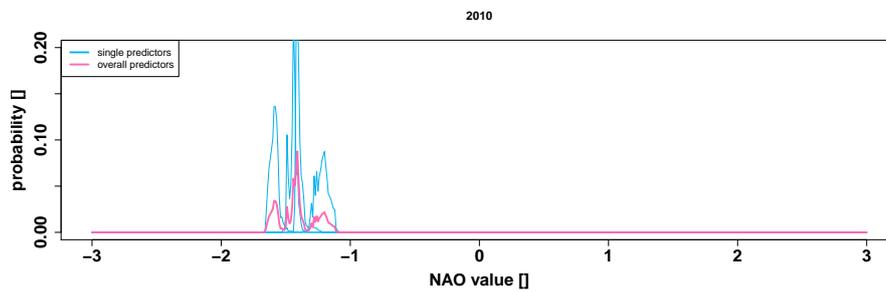
Figure 3. Predictor prediction of the NAO for 2010. Single predictors (light blue) as pdf of their bootstrapped uncertainties. From this the overall predictor prediction (pink) is created by empirical mixture modelling.
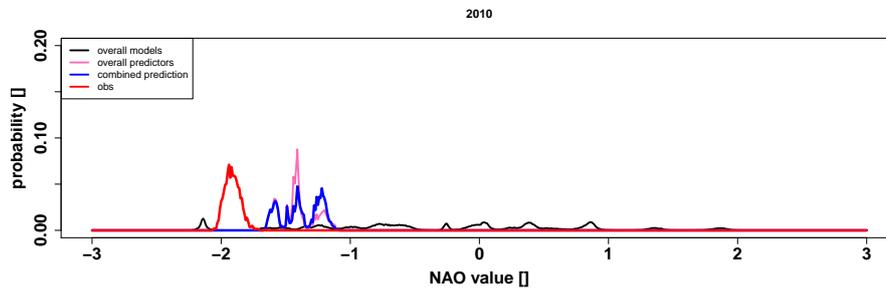
Figure 4. Sequence of post-processing procedure for the NAO in 2010. Combining overall models (black) and overall predictors (pink) to a combined prediction (blue) and comparing it to the observations (red).
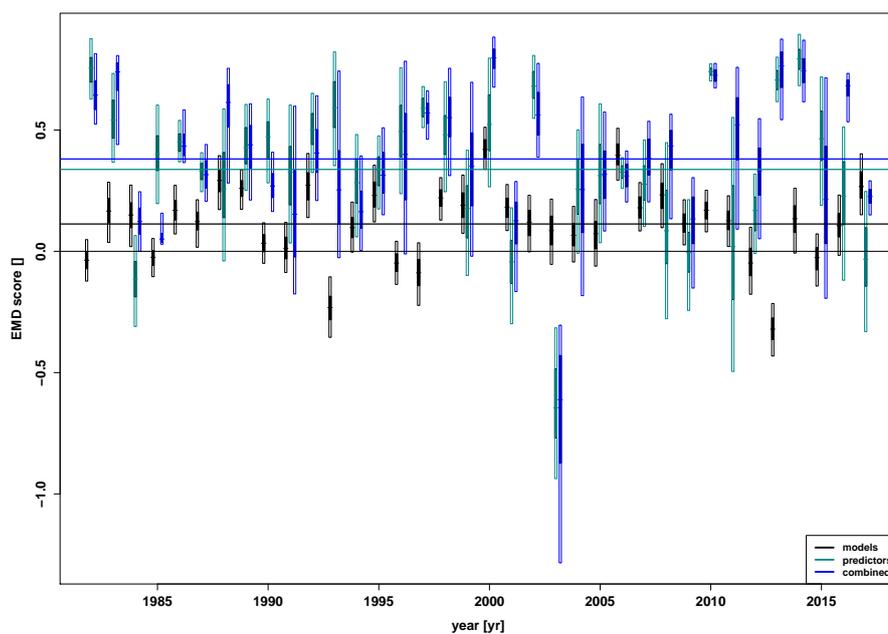
Figure 5. Yearly comparison of the WNAO scores for overall model (black), overall predictors (dark green) and combined prediction (blue). Each vertical bar represents the 5% to 95% bootstrapped 1D-continuous-EMD score. The filled part of these bars are the 25th to 75th quartiles and the small vertical lines the associated median. The long vertical lines are the averaged yearly scores for the different predictions.
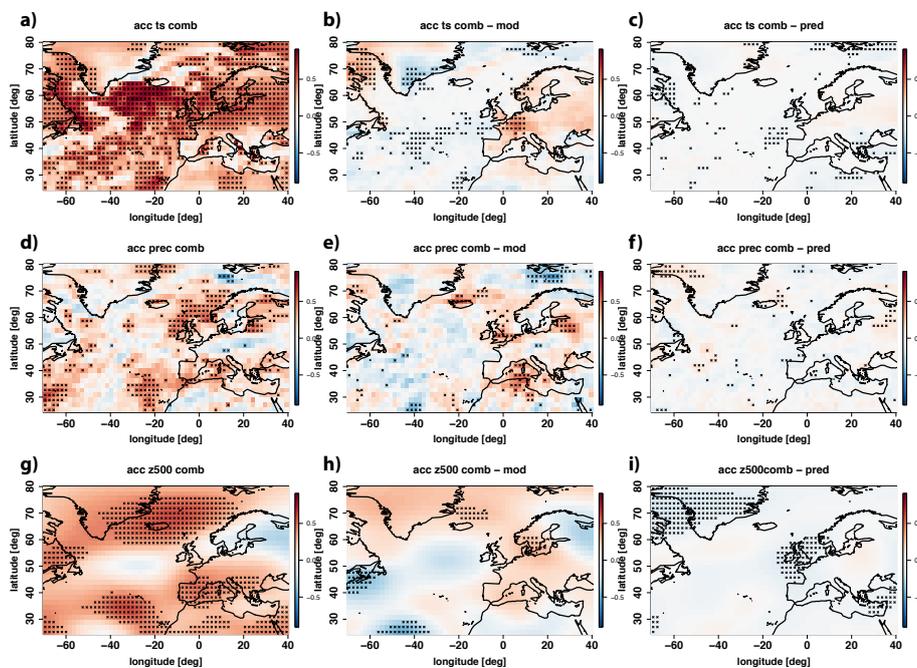
Figure 6. ACC results for the WNAO for three different atmospheric variables: surface temperature (first row), total precipitation (second row) and geopotential height (third row). Shown are the combined prediction (first column), the difference between the combined and the model prediction (second column) and the difference between the combined and the predictor prediction (third column). Black dots indicate significances estimated by a 500 sample-bootstrap.
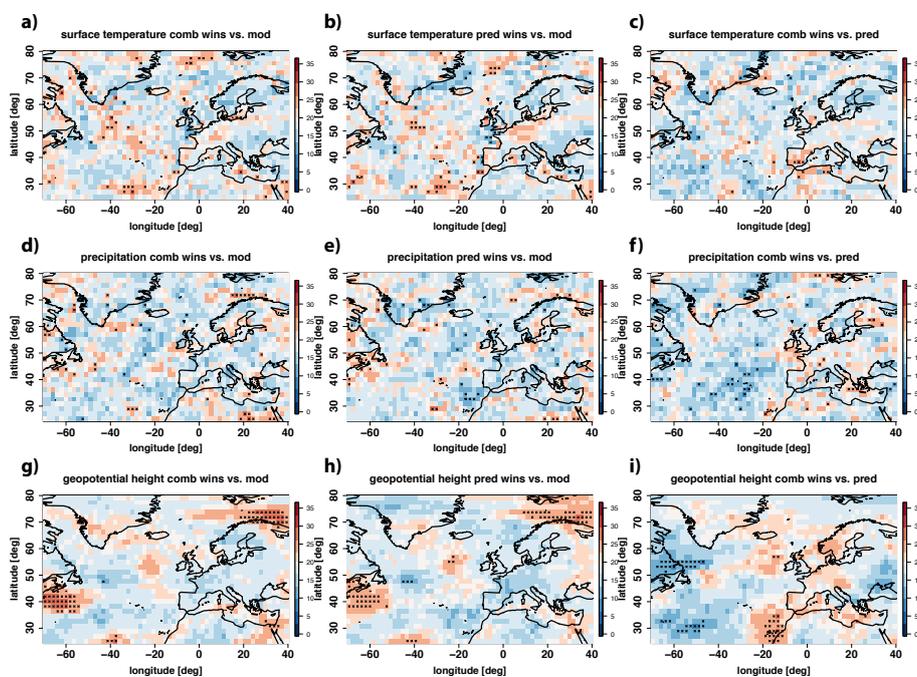
Figure 7. Relative positioning of the predictions of the variable fields on the basis of the 1D-continuous-EMD score. Shown is the number of years in which the first named prediction has a better score than the second named. Compared are the combined with the model prediction (first column), predictor with the model prediction (second column) and combined with the predictor prediction (third column). Significances are determined by a comparison towards a random walk at a confidence level of 0.05. Variables are positioned as in figure 6.